AFRL-AFOSR-UK-TR-2015-0001

SATA: Stochastic Algebraic Topology and Applications

Professor Robert Adler

TECHNION R & D FOUNDATION LTD
SENATE BUILDING ROOM 157 - TECHNION CITY
HAIFA, ISRAEL

EOARD GRANT FA8655-11-1-3039

Report Date: September 2012

Final Report from 1 September 2011 to 31 August 2012

Air Force Research Laboratory
Air Force Office of Scientific Research
European Office of Aerospace Research and Development
Unit 4515, APO AE 09421-4515

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE ***(DD-MM-YYYY)*** 01-09-2012	**2. REPORT TYPE** Final Report	**3. DATES COVERED** *(From – To)* 01 Sep 2011 – 31 Aug 2012

4. TITLE AND SUBTITLE **SATA: Stochastic Algebraic Topology and Applications**	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER **FA8655-11-1-3039**
	5c. PROGRAM ELEMENT NUMBER 61102F
6. AUTHOR(S) **Prof Robert Adler**	**5d. PROJECT NUMBER**
	5d. TASK NUMBER
	5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) TECHNION R & D FOUNDATION LTD SENATE BUILDING ROOM 157 - TECHNION CITY HAIFA, ISRAEL	**8. PERFORMING ORGANIZATION REPORT NUMBER** N/A
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) EOARD Unit 4515 APO AE 09421-4515	**10. SPONSOR/MONITOR'S ACRONYM(S)** AFRL/AFOSR/IOE (EOARD)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) **AFRL-AFOSR-UK-TR-2015-0001**

12. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report covers the first 3 years of a planned 5 year research project. The final 2 years is funded under grant FA9550-15-1-0025.
The Stochastic Algebraic Topology and Applications (SATA) project aims to exploit recent advances in the complementary areas of topology and stochastic processes to tackle a wide range of data analytic problems of broad importance. Treating data topologically is crucial in scenarios in which it is important to detect, localize, and perhaps perform an initial classification of objects without attempting to completely characterize them. Adding a stochastic element allows for the almost pervasive situation in which the data itself is imperfectly observed due to the presence of background noise. As current probabilistic and statistical methodology is ill suited to detect such qualitative structures, the project aims to develop generic stochastic models whose topological structures are amenable to mathematical analysis, as a first step towards implementation of a broader, more quantitative program. Core topics include random functions on manifolds, random manifolds created by random embeddings, and random manifolds arising in machine learning, along with their theoretical and practical interplay. Secondary topics include the analysis of associated algorithms, and the topological understanding of random spaces that arise in particular stochastic models. We have also studied implementation and application of these ideas on some problems coming from engineering and physics. Initial results were obtained regarding the statistics of random functions, the application and analysis of Morse theory in random settings, and on the complexity of the basic topological inference problems in data analysis. Significant progress has been made in the areas of the Statistics of Random Functions; Morse Theory, Critical Points, Betti Numbers and Random Complexes; and Random Manifolds and Random Embeddings.

15. SUBJECT TERMS

EOARD, Network Theory, Sensor Technology, Mathematical Modeling

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18, NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** James H Lawton, PhD
a. REPORT UNCLAS	**b. ABSTRACT** UNCLAS	**c. THIS PAGE** UNCLAS	SAR	16	**19b. TELEPHONE NUMBER** *(Include area code)* (703)696-5999

AFOSR New Award # FA9550-11-1-0216 and Combines

Previous EOARD Grant# FA8655-11-1-3039

Stochastic Algebraic Topology and Applications

Performance Report for the Period (9/2011-9/2013)

PI.
Shmuel Weinberger
Department of Mathematics
University of Chicago

CoPIs

Yuliy Baryshnikov
Departments of Mathematics and Electrical and Computer Engineering
University of Illinois at Urbana-Champaign

And

Jonathan Taylor
Department of Statistics
Stanford University

Summary

This report summarizes the second year of a five year project involving three US investigators together with Robert Adler and his team, at the Technion (Israel) who is working on a similar grant with the European Office of Aerospace Research and Development. We have not always been able to separate the work done by the American group from the Israeli team.

Progress over the first years has been considerable. Work has begun in all four locations with a group that includes the PI's, graduate students, some postdocs and other collaborators. Work has begun on about half of the specific projects presented in the original proposal, and results have already been obtained on a number of these. These include results related to the statistics of random functions, random complexes, and random manifolds and embeddings. In addition, already there have been some extensions of the scope of the project to include some new problems and areas related to the theme (stochastic algebraic topology) that were not mentioned in the original proposal.

There has also been substantial training of graduate students and postdoctoral fellows, dissemination of results and general educational activity on the importance of stochastic algebraic topology as tool in data analysis.

Objectives.

Recall that the Stochastic Algebraic Topology and Applications (SATA) project aims to exploit recent advances in the complementary areas of topology and stochastic processes to tackle a wide range of data analytic problems of broad importance. Treating data topologically is crucial in scenarios in which it is important to detect, localize, and perhaps perform an initial classification of objects without attempting to completely characterize them. Adding a stochastic element allows for the almost pervasive situation in which the data itself is imperfectly observed due to the presence of background noise. As current probabilistic and statistical methodology is ill suited to detect such qualitative structures, the project aims to develop generic stochastic models whose topological structures are amenable to mathematical analysis, as a first step towards implementation of a broader, more quantitative program. Core topics include random functions on manifolds, random manifolds created by random embeddings, and random manifolds arising in machine learning, along with their theoretical and practical interplay. Secondary topics include the analysis of associated algorithms, and the topological understanding of random spaces that arise in particular stochastic models. We have also studied implementation and application of these ideas on some problems coming from engineering and physics.

Status of Effort

This is basically a mathematics-based project, and the methods are hard analysis supplemented with, and often motivated by, computation. The core ``raw material" for this project is therefore appropriate manpower, which we have been successful in drafting at the level of graduate students and postdocs. Baryshnikov and Weinberger have visited the Technion during the past twelve months, and all of three investigators have done so in the previous years. They have also met in pairs at Chicago and Stanford. The researchers have discussed, thought about individually and then in groups, key issues related to the proposal. When appropriate, pieces have been broken off and discussed with students and postdocs or other outside collaborators. Our ideas are tested by example, the search for proof, and numerical simulation.

Initial results were obtained regarding the statistics of random functions, the application and analysis of Morse theory in random settings, and on the complexity of the basic topological inference problems in data analysis, as explained in more detail below.

Accomplishments/New Results

Below we describe three general areas in which we have made significant progress and a fourth area of new ideas that are spurred by our projects. The headings all correspond to topics in the original proposal, which provides background material.

1: Statistics of Random Functions

From a mathematical standpoint, our basic setup here starts with a base topological space, M, typically but not always a manifold, and a random function f on M, with values in $\mathbf{R}^n$ for some n≥ 1. Given a realization of such a function, we proposed studying various statistics pertaining to it, including properties of the set of critical points, critical values, sub- and super-level sets, etc.

Several developments have been made in this area. The first lies in the study of excursion sets, and, in particular, the topological issue of whether or not two or more points which lie in an excursion set belong to the same connected component, a basic question that has constantly eluded analysis.

In joint work with Gennady Samorodnitsky (Cornell), Adler studied a one-dimensional version of this problem from the point of view of large deviations, finding the asymptotic probabilities that two such points are connected by a path lying within an excursion set, and so belong to the same component. In addition, we obtain a characterization and descriptions of the most likely paths, given that one exists. The resulting paper [2] represents somewhat of a breakthrough in terms of tools for studying the topology of Gaussian random fields in all dimensions, and plans to continue this investigation over the coming years.

The issues involving components are related to the "Betti-0 Persistence Homology" of the (graph of the function). The most mysterious part of persistence homology is the meaning or interpretation of short "bars". One can show that the presence of very many small bars (on a smooth manifold) indicates large norm in various function spaces: in particular, Holder exponents give considerable control on the number of small bars of length t. In an exploratory study of this in the highly non-smooth situation of 1 dimensional Brownian motion (and Brownian bridge). One can express the expected number of PH intervals overlapping a specific value of the parameters a<b in terms of certain theta functions using a version of the reflection principle,. We (i.e. the whole group) thus established two power laws for the length t intervals: their aggregate number grows like $1/t^2$. The number of such intervals that pass through a particular level (of the function) grows like 1/t. This was done last year, but we have not yet had the chance to write up these results for publication; partly this is because later work seems more pressing and is changing our points of view on these results. In particular, this work has been impacted by the study of various metrics and measures on spaces of barcodes. In addition our point of view on these questions has also been changed by some later developments on the Brownian Bridge: For a Brownian bridge, the persistence diagram (or, rather, the point process, placing a point mass at (b; d) $\in R^2$ for each (b, d) bar in PH0), has nontrivial fractal patterns. Baryshnikov found the intensity density for the persistence point process (rep-
resentable as a certain mock theta function). The density diverges as $(d\text{-}b)^3$ near the diagonal.
Baryshnikov with a group of graduate students started investigations of the higher cumulants of the persistence point process. It turned out to be related to understanding the Moebius functions on a sequence of (infinite) posets, and leads to a family of theta-like functions.

In joint work of Adler and Taylor with Eliran Subag, a Technion graduate student, they provided a new approach, along with extensions, to results in two important papers of Worsley, Siegmund and coworkers closely tied to the statistical analysis of fMRI (functional magnetic resonance imaging) brain data. These papers studied approximations for the exceedence probabilities of scale and rotation space random fields, the latter playing an important role in the statistical analysis of fMRI data. The techniques used there came either from the Euler characteristic heuristic or via tube formulae, and to a large extent were carefully attuned to the specific examples of the paper. In [1] they treated the same problem, but via calculations based on the so-called Gaussian kinematic formula. This allowed for extensions of the Worsley-Siegmund results to a wide class of non-Gaussian cases. In addition, it allows one to obtain results for rotation space random fields in any dimension via reasonably straightforward Riemannian geometric calculations. Previously only the two-dimensional case could be covered, and then only via computer algebra.

A more applied project which aims to take advantage of new results in applied topology related to the Euler integrals of Robert Ghrist and Yuliy Baryshnikov, was started over the past year. The underlying idea, which also exploits theoretical results of Borman and Bobrowski, is to compute the Euler integrals of `signal + noise', and to use the resulting integral as a statistic to test for the presence of the signal. This project is still at the stage of experimental testing to see whether the ideas work or not in practice, and the initial results are encouraging. At the same time, work is progressing on central limit theorems for Euler integrals of random processes observed over a long time or large space.

One of the questions that we ultimately hope to answer through the deeper analysis of the structure of these fields relates to the depth of local minima, which would give useful information on the distribution of 0-persistence intervals of sub-level sets in higher dimensional and smooth situations.

Nonlinear Theory

A first essentially nonlinear problem consists of understanding maps from the circle into a Riemannian manifold. One of the frequent question in applied topology is the search for the minimal (in some sense) representations of geometric structures under some topological constraints. An archetypal example is the search for the shortest loop representing a given (free) homotopy class on a Riemannian manifold. Baryshnikov, with Y. Mileyko and M. Arnold approached this question from the stochastic viewpoint, arguing that a randomly sampled one enough discretized loop in a given class becomes close to the optimal loop, as the mesh size decreases. Let M be a compact Riemannian manifold, and that γ is a free homotopy class. Assume that representatives of γ of length less than L exist. We replace a loop (i.e. mapping $S^1 \rightarrow M$) with a discrete loop, i.e. a sequence x_0 , x_1 ,… $x_N = x_0$ such that the distances are bounded by d_M(xi, xi+1) <L/N. When N is large enough, L/N is less than the injectivity radius of M, and we can complete the discretized loop [x] = x0; : : : ;xN to a continuous loop lx : $S^1 \rightarrow M$. We denote the set of the discretized loops lx such that the free homotopy class of lx equals γ by $\Lambda(L;N;\gamma)$. It is an open subset of M^N, inheriting Lebesgue measure from M^N. Let $L*(\gamma) = \inf L(s)$, the infimum of the lengths of the continuous loops s : $S^1 \rightarrow M$ representing γ. Arnold, Baryshnikov and Mileyko proved that as $N \rightarrow \infty$, the randomly sampled from $\Lambda(L;N;\gamma))$ have lengths close to $L*(\gamma)$: The probability that a random curve has length > $L+\varepsilon$ is bounded by $A\exp(-N\psi(L-L*(\gamma))/\varepsilon)$ for a rate function ψ on $\Lambda(L;N;\gamma))$ and with high probability, the curve is actually close to an optimal loop. In other words, for fine enough loops, the optimization may be achieved by drawing a random discretized loop from the set , and running a MCMC algorithm (e.g., Gibbs sampling) for a while.

The question of how long this would need to run, and what N should be, is governed by aspects of the fundamental group itself, as is implicit in the paper of Weinberger [10] that shows that the non-deterministic Morse landscape of the energy functional is controlled by the Dehn function of the fundamental group (and largely independent of the Riemannian metric on the manifold because of the stability properties of persistent homology).

2: Morse Theory, Critical Points, Betti Numbers and Random Complexes

As described in the original proposal, we planned to invest considerable effort in the study of topological properties of random simplicial complexes. This has led to a number of new results.

Omer Bobrowski and Adler [3] considered a finite set of points P in $\mathbf{R}^{\mathbf{d}}$ and the ``distance function" d_P: $\mathbf{R}^{\mathbf{d}} \rightarrow \mathrm{R}$ which measures Euclidean distance to the set P. They studied the number of critical points of d_P when P is a random sample from a given distribution. In particular, we studied the limit behavior of N_k -- the number of critical points of d_P with Morse index k-- as the number of points in P goes to infinity.

They found explicit computations for the normalized, limiting, expectations and variances of the N_k as well as distributional limit theorems. These results are related to recent results of Matt Kahle in which the Betti numbers of the random Cech complex based on P were studied and these ideas are relevant to the work on the persistent homology of noise discussed below.

A different problem where the Morse theory of distance functions is applied is the problem of inferring topological invariants from point cloud data. Weinberger's student, Katharine Turner, [4], dealt with the situation of complicated subsets of manifolds with bounded geometry, and gave conditions for when a Niyogi-Smale-Weinberger type algorithm would work in this more complicated setting (as the subset is not smooth, and the background is non-Euclidean --- both important weakenings of the hypotheses of the original results on this problem).

The practical implication of these results lies in the possible design of sampling algorithms for manifold or more general topological learning, with significance tests, via approximating simplicial complexes.

Adler, Bobrowski and Weinberger made considerable progress is the analysis of the persistent homology of pure noise. In particular, they studied a phenomenon that they have named ``crackle" wherein at every scale homology classes form only to die at larger scales (or with the addition of more data). Once again, random Cech complexes are created (the Rips theory is similar in outline, although different in detail), this time based on fixed inter-point distances, but based over different types of samples. The three examples we studied involved sampling with Gaussian, exponential, and heavy tailed noise.

They found remarkably different behavior of the topology of the complexes in the three cases. In particular, they discovered phenomena that have serious implications for manifold learning using purely topological techniques, since spurious homological ``crackle" appears. For example, while estimating the homological structure of low dimensional manifolds from high dimensional data in the presence of Gaussian (or other low tailed) noise seems to work quite well, this will not be the case for heavier tailed noise. It thus seems quite possible that the work of Niyogi-Smale-Weinberger on topological learning in the presence of noise, wherein ambient dimensional independent estimates were possible in the Gaussian case cannot be extended even to the situation of exponentially decaying noise. This phenomenon has been observed experimentally in the past, but we are now developing a theoretical understanding, which is necessary to provide more effective algorithmic procedures for manifold learning in problematic scenarios.

Continuing in this direction, the limit time behavior, for dense sampling of Riemannian manifolds, was studied by similar methods. Baryshnikov and Weinberger had observed that if V denotes the ratio of the volume of a manifold to the volume of the Euclidean ε-ball, then while the number of components is correct when we compute Cech homology using N = 2^{-d} VlogV points, and all of the homology is correct at the scale N = VlogV. While we had believed that the homology groups in dimensions between 1 and N-1 all occur at different fractions of VlogV, the crackle technology applied to this problem shows that this is not the case: they all are computed accurately at the VlogV scale - more or less simultaneously.

Further work in progress at the VloglogV scale seems to indicate a finer structure to the convergence of Betti numbers (and that lower Betti numbers do converge slightly earlier). We have also investigated manifolds with boundary and certain stratified spaces that show new phenomena because of singularities.

Finally (for this direction), in [9], lower complexity bounds were obtained for several types of complexity (such as Kolmogorov, a.k.a. description complexity, sample complexity, and decidability, as precursor to computational complexity) of basic topological problems including dimension determination, topological type, singularitiy detection. These were considered in both noiseless and noisy environments.
The tools used were a mix of topological, logical, and probabilistic methods. This paper has already appeared, and is somewhat improved over the version submitted last year.

The lower bounds in these problems pose an important issue for TDA: many of the usual questions people ask are unfeasible in general: computation of invariants is too difficult, the number of topological types is too large. Applications of topological methods must either explain why the data should be suitable for those methods - e.g. why the complications that could arise, do not -- or they should be focused on invariants that can be measured. Weinberger has begun a study of such invariants, modeled on testability properties of graph properties. The simplest of these is the Euler characteristic divided by the volume - which is essentially (for Riemannian manifolds) the average (Pfaffian of the) curvature. As an average, it is subject to sampling. Thus, a large submanifold in Euclidean space whose average curvature is large will surely have complicated topology, and discovering its topological properties will require enormous sampling and computational resources.

Similarly, characteristic 0 Betti numbers seem to be testable (but not too straightforwardly: random regular graphs have high Euler characteristic and first Betti number: but randomly they look like trees that have no local topology!). Whether this is the case for mod p Betti numbers is an important problem. The logarithm of the p-torsion in the homology has been shown to **not** be testable, even when normalized by volume.

These ideas relate also to the theory of quasicrystals and statistical physics (mentioned last year and below), and has connections to some current themes in number theory (and the theory of lattices in Lie groups).

K.Turner [5] has been working with the group at Duke to develop tools for studying "average" persistence diagrams to facilitate the import of statistical tools.

One of the results about "crackle" that will appear in a joint paper of all the members of the team is that for 1 dimensional exponential distributions the following formula (with a suitable law of large numbers) gives the average (with large number of data points) PH_0

$$E(PH_0(\text{N points chosen from an exponential distribution})) \to \int e^{-x}/(1-e^{-x})^2 \, [0, x/2] \, dx$$

Note the strange nature of the integrand (note that intervals appear in it![1]): it is essentially a current describing the average number of times one should see an interval approximately of length [0, x/2]. The blow up near the origin is because, with many points, there are very many very short persistence intervals. Its quadratic nature is perhaps typical. Similar things occur for Brownian motion, as mentioned above. The nonzero measure associated to long intervals is essentially a precise form of the crackle phenomenon.

3: Random Manifolds and Random Embeddings

In the initial proposal, we noted that random manifolds arise in a number of scenarios, and that one of the key geometric quantities that arises there in recovering the homology of a manifold $M \subset \mathbf{R}^n$ by randomly sampling points from it is the critical radius, a.k.a. the feature size, or the distance to the medial axis, τ, of the manifold. This quantity depends on the embedding and thus it is of interest to study the behavior of the critical radius of a Riemannian manifold (M,g) for a generic, or random embedding of M into $\mathbf{R}^n$ for large n. This is of importance in determining the extent to which generically the phenomenon of identifying a manifold from Gaussian noisy data depends on the dimension of the manifold and to what extent it depends on the ambient dimension that the data points lie in.

[1] Some might prefer the intervals replaced by characteristic functions of the intervals, and then this formula can be viewed as an equality in a space of measures.

A natural model to consider is based on taking independent, identically distributed, copies, $f_1 \ldots f_n$ of a real-valued random field on M and then working with f = $(f_1 \ldots f_n) : M \rightarrow \mathbf{R}^{\mathbf{n}}$ to define the random, embedded manifold f(M).

Each such random embedding gives rise to a random Riemannian metric g_f on M, that is naturally related to the original metric g. Other geometric features of interest include the study of the geometric invariants of such Riemannian metrics.

Adler, Taylor, and Weinberger have made considerable progress on studying these embeddings, but some estimates currently seem somewhat more subtle than we thought at first. They have also found a way to exploit random embeddings to estimate underlying geometric (i.e. not just topological) properties (such as curvature integrals) of learnt manifolds, using purely topological data. We hope to complete the proof that a random embedding of a Riemannian manifold is a learnable subset of Euclidean space with sample complexity independent of ambient dimension, as well as develop tools for the estimation of its geometric properties.

4. Other directions that have grown out of this work.

Taylor has written a paper on a significance test for the LASSO. [7] This paper describes a testing procedure that can be used to decided when to stop regularizing in a machine learning algorithm such as the LASSO. The test explicitly adapts to the searching through model space, hence its null distribution is properly calibrated, unlike the usual procedures such as forward stepwise. This paper has offered new insight into something as established as forward stepwise regression, which is ongoing work with a student (Joshua Loftus, Ph. D. Stanford). The approach also generalizes to arbitrary seminorm penalties such as the group LASSO, nuclear norm, etc. The main tools used in this generalization is based on the theory of smooth random fields developed by PIs Adler and Taylor. The tools can also be used to establish results on smooth random fields, particularly the correct generalization of the superexponential accuracy of the expected Euler characteristic heuristic for Gaussian random fields that are not marginally stationary.

His second paper in this series is “Adaptive testing for the graphical lasso” This paper is another application of the approach used in the previous paper. In this paper, he applied this sequential testing procedure to the graphical LASSO algorithm. Roughly speaking, the results identifies a
(well-calibrated) test that can be used to correctly identify the connected components formed by the precision matrix of a Gaussian random vector, under the assumption that the population precision matrix is sparse.
This same procedure can also be used to determine at what point in a single-linkage clustering algorithm to stop merging clusters when the pseudometric used in the clustering algorithm is based on a correlation matrix.

Baryshnikov, and Weinberger with Turner have been studying related problems on networks. The persistence homology of networks (with analytic growth bounds, with

respect to steadily increasing scale) has been related to network flow and congestion problems. dimensions of networks. Baryshnikov and his postdoc (now assistant professor at U Hawaii) Yuri Mileyko continued the studies of the (local) dimensions of synthetic and real-life networks. The background for this quest was an emerging trend in networking community to view networks as hyperbolic in some sense, or other (say, being CAT(0) spaces, or Gromov-hyperbolic etc). One thread in this area of research had as an underlying premise that the real-life networks can be properly modeled by Random Geometric Graphs sampled from a ball in homogeneous hyperbolic spaces, or, even more specifically, from the hyperbolic plane. While the pictorial representations in the numerous publications in this spirit felt convincing, the basic questions were not asked, i.e. why plane? Why the assumptions of homogeneity? etc.
In general, the random finite metric space obtained a by dense enough sampling from a Riemannian manifold would provide enough data to detect at least the dimension of the underlying manifold: if X ⊂ Mm is a finite sample from M, then for spherical shells of points in X, and judiciously chosen radii R; r (R much less than the injectivity radius, r large to ensure dense sampling), the persisting homology of the Rips complex of SX(x; R; r) should be concentrated in dimensions 0 and m, for interior points of the sample, and just in dimension 0 for the points near the boundary. Experiments confirmed that this is exactly what happens for the samples from hyperbolic plane. However, contrary to what one might expect from the existing literature, the analysis of the ASN network (the world-visible structure of the autonomous domains, roughly the network of Internet connections) shows that their local homologies are extremely wild and irregular, and are nowhere close to the sample from the hyperbolic plane (or any manifold). On the positive side, the local homology is yet another characteristic of the nodes in large graphs, and we plan to use it systematically for network analysis (and, perhaps, to analyze samples from singular spaces, in a TDA fashion). The results of the experiments are visible by invitation at a web site that will
go public in the near future.

Weinberger has been working with Bellissard and Ulgen-Yildirim on problems of material science that have traditionally been approached using methods of statistical physics by methods related to the Novikov conjecture in non-commutative geometry. So far, these have not led to anything new about material sciences, but the types of limits that arise and are studied analytically are a counterpart to the measure theoretic limits occurring in Banjamini-Schramm convergence theory. This is an important direction for TDA purposes as explained above: it leads to a theory of which invariants are computable and which are not -- or alternatively it potentially will lead to tests that can reject the "simple geometry + noise hypothesis". One such test is very similar in spirit to the the Baryshnikov-Mileyko test, but there are others not based on the manifold hypothesis.

With S. Mukherjee, Katharine Turner has been working on using persistent homology in shape statistics – describing a shape by a persistent homology transform considering the height functions is all the different directions. They have been working with D. Boyer in the Evolutionary Anthropology department at Duke University on actual data and have obtained promising results comparing calcanei bones of various primates, using the theory of means of bar codes.

Finally, we mention work of Baryshnikov on understanding the Battery: Consider a planar domain B (say, a rectangle) with a selected ark $A \subset \partial B$, and a collection of non-overlapping disks $Bl \subset B$; l = 1, 2, …n inside it. These data define a mapping from the configuration space of the non-overlapping displacements of the disks Bl into the moduli space of real hyperelliptic curves of genus 2m: to each placement of the particles one can associate a hyperelliptic curve C (so that the configurations corresponding to the particles approaching each other or the boundary of B are mapped to the boundary of the moduli space. The significance of this mapping comes from the problem of optimization of the anode placements in lithium-ion batteries. The particles accumulating the ions can be modeled as the disks Bl placed in a electrolyte container B. The figures of merit in this model is a certain period on the hyperelliptic surface C corresponding to the charging flows to the anode particles. These flows should be maximized and equilibrated. In essence, the flows correspond to the derivatives of the Dirichlet kernel with respect to the normal parameters on the boundary. This problem has been addressed by Baryshnikov together with Anil Hirani and his graduate student Kaushik Kalyanaraman. Code for performing computations using Discrete Exterior Calculus was written, debugged and tested and is being used now in the optimization procedures.

Personnel Supported

Shmuel Weinberger, PI,
Department of Mathematics,
University of Chicago

Yuliy Baryshnikov, co-PI
Departments of Mathematics and Electrical and Computer Engineering
University of Illinois at Urbana-Champaign

Jonathan Taylor, co-PI
Department of Statistics
Stanford University

Partially supported the related joint work of and with:

Robert Adler[2]
Department of Electrical Engineering
Technion

Katharine Turner (Graduate Student)
Mathematics
University of Chicago

Han Wang (Graduate Student)

[2] Adler's EOAR grant supported additional postdocs and students in Israel.

Mathematics
University of Illinois at Urbana-Champaign

Publications

[1] R.J.Adler, E.Subag and J.E.Taylor, Rotation and scale space random fields and the Gaussian kinematic formula, Annals of Statistics, 40, 2012, 2910-2942.

[2] R.J.Adler, E.Moldavskaya and G.Samorodnitsky, On the existence of paths between two points in high level excursion sets of Gaussian random fields. (Annals of Probability, to appear)

[3] O.Bobrowski and R.J.Adler, Distance functions, critical points, and topology for some random complexes. Annals of Applied Probability, 2013. Under revision, waiting for response from the Journal

[4] K.Turner, Cone fields and topological sampling in manifolds with bounded curvature, Journal of FoCM (to appear)

[5] K.Turner, Y.Mileyko, S.Mukherjee, J.Harer, Fréchet Means for Distributions of Persistence diagrams (submitted for publication)

[6] K.Turner, E. Munch, P. Bendich, S. Mukherjee, J. Mattingly, J. Harer, Probabilistic Fréchet Means and Statistics on Vineyards (submitted)

[7] J.Taylor, A Significance Test for the LASSO, (preprint)

[8] J.Taylor, Adaptive testing for the graphical LASSO (preprint)

[9] S.Weinberger, The complexity of some basic topological inference problems, Journal of FoCM, (to appear, available online)

[10] S.Weinberger, What is…Persistent Homology? Notices AMS January 2011 pp. 36-39

Interactions/Transitions

All the PIs and co-PIs have met in either Stanford, Chicago, and Haifa over the past year.

In addition to the basic research described above, we have been involved in dissemination of the basic ideas of SATA at a number of venues.

Adler is coordinating a tutorial on An Introduction to Statistics and Probability for Topologists at the IMA in October 2013, as well as being one of the organizers of a workshop on Topological Data Anaysis which will follow the tutorial session.

Taylor spoke at a special Topological Data Analysis workshop at NIPS in December 2012. He is an invited speaker at the European Meeting of Statisticians and will participate at the IMA workshop in October 2013.

Baryshnikov and Weinberger both gave talks related to SATA at conferences:

Weinberger gave a plenary talk at the Applied Algebraic Topology meeting in Bedlewo. He gave the "Frontiers of Mathematics" lecture series at Texas A&M; one of the lectures featuring ideas related to property testing and its connections to both pure and applied problems. He will be lecturing at two or three workshops at IMA during the coming year.

Baryshnikov presented some of the results on random networks at NIST-Bell Labs workshop on Geometry of Networks, at NIST, the MCA special session on Applied Algebraic Topology, and a plenary talk at the SIAM conference on Applied Algebraic Geometry.

Turner gave the following talks:
Fréchet means of persistence diagrams, Data Seminar, Duke University (February 2013)
Fréchet means of persistence diagrams, AMS Fall Central Sectional Meeting (Special Session on Applied Topology), (October 2012)

Statistics of persistence diagrams and Shape statistics via persistence homology, Topological Data Analysis Workshop, Defence Science Institute, Melbourne (June 2013)
Shape statistics via the persistent homology transform, Australian National University (August 2013)

Training of Graduate Students and Postdoctoral Fellows

The past twelve months have been very active in terms of graduate students and postdoctoral fellows. In Israel, Omer Bobrowski finshed his PhD thesis on The Algebraic Topology of Random Fields and Complexes and has taken up a research assistant professorship at Duke in August. He has continued working closely with Adler and Weinberger, and has started collaborations with the Duke group (e.g. Murkhejee).

Weinberger's student, Katharine Turner, has done work related totopological inference and developing statistical models for persistent homology (mentioned above), the latter with the Duke group that Bobrowski has joined.

Eliran Subag completed his MSc thesis on Rotation and Scale Space Random Fields and Mixing Times for Random Shuffles the major part of which was directly related to SATA.

Two new graduate students, Yonatan Rosmarin and Gregory Naitzat began their research projects with Adler. Naitzat is studying the use of Euler integration as a tool in signal and image analysis, and Rosmarin is studying perturbation theory as a tool for studying the topological properties of sets generated by non-Gaussian random processes. Yogeshwaran Dhandapani has been working as a postdoctoral fellow for SATA for almost two years. He has already obtained nice results on the homology of random simplices built over random point sets which exhibit dependence. As mentioned above, the behavior of the corresponding complexes exhibits new scaling phenomena and this has signi_cant implications for the (non) robustness of applied topological methods based on implicit assumptions of an underlying independence. He is currently working on similar problems for what is known as the `thermodynamic', or `percolation' range for random complexes.

While not supported under SATA, another postdoctoral fellow, Anthea Monod, has been working on related topics, and in Fall 2013 two new postdocs and one PhD student will be joining the overall applied topology effort at the Technion.

Han Wang, a PhD student at UIUC started recently his work under Baryshnikov's supervision.

Honors/Awards

Adler was invited to give a plenary Special Invited Lecture at the European Meeting of Statisticians in Budapest, July 2013 and was awarded a prestigious European Research Council Advanced grant.

Weinberger is among the inaugural class of AMS Fellows announced in November 2012.